(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 January 2003 (23.01.2003)

PCT

(10) International Publication Number
WO 03/007235 A1

(51) International Patent Classification[7]: G06N 5/02

(21) International Application Number: PCT/US02/22460

(22) International Filing Date: 11 July 2002 (11.07.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/304,647 10 July 2001 (10.07.2001) US
10/192,783 9 July 2002 (09.07.2002) US

(71) Applicant: AUDIBLE MAGIC CORPORATION [US/US]; 985 University Avenue, Suite 35, Los Gatos, CA 95032 (US).

(72) Inventor: WOLD, Erling, H.; 5618 Ludwig Avenue, El Cerrito, CA 94530 (US).

(74) Agent: SIERRA PATENT GROUP, LTD.; P.O. Box 6149, Stateline, NV 89449 (US).

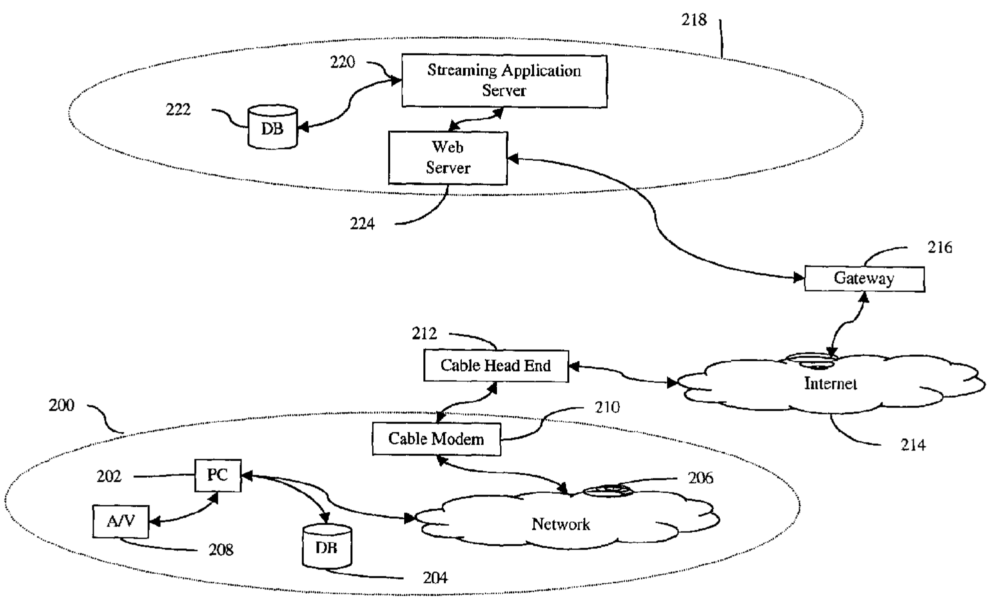
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD AND APPARATUS FOR IDENTIFYING AN UNKNOWN WORK



Present Invention

(57) Abstract: A method and apparatus for identifying an unknown work is disclosed. In one aspect, a method may include the acts of providing a reference database (204) having a reduced dimensionality (106) containing signatures of sampled works; receiving a sampled work; producing a signature from the work; and reducing the dimensionality of the signature (100).

WO 03/007235 A1

METHOD AND APPARATUS FOR IDENTIFYING AN UNKNOWN WORK

PRIORITY CLAIM

This application claims the benefit of United States Provisional Application Serial No. 60/304,647, filed July 10, 2001.

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to data communications. In particular, the present invention relates to a novel method and apparatus for identifying an unknown work.

The Prior Art

Background

Digital audio technology has greatly changed the landscape of music and entertainment. Rapid increases in computing power coupled with decreases in cost have made it possible for individuals to generate finished products having a quality once available only in a major studio. One consequence of modern technology is that legacy media storage standards, such as reel-to-reel tapes, are being rapidly replaced by digital storage media, such as the Digital Versatile Disk (DVD), and Digital Audio Tape (DAT). Additionally, with higher capacity hard drives standard on most personal computers, home users may now store digital files such as audio or video tracks on their home computers.

Furthermore, the Internet has generated much excitement, particularly among those who see the Internet as an opportunity to develop new avenues for artistic expression and communication. The Internet has become a virtual gallery, where artists may post their works on a Web page. Once posted, the works may be viewed by anyone having access to the Internet.

One application of the Internet that has received considerable attention is the ability to transmit recorded music over the Internet. Once music has been digitally encoded, the audio may be both downloaded by users for play, or broadcast ("streamed") over the Internet. When audio is streamed, it may be listened to by Internet users in a manner much like traditional radio stations.

Given the widespread use of digital media, digital audio files, or digital video files containing audio information, may need to be identified. The need for identification of digital files may arise in a variety of situations. For example, an artist may wish to verify royalty payments or generate their own Arbitron®-like ratings by identifying how often their works are being streamed or downloaded. Additionally, users may wish to identify a particular work. The prior art has made efforts to create methods for identifying digital audio works.

However, systems of the prior art suffer from certain disadvantages. One area of difficulty arises when a large number of reference signatures must be compared to an unknown audio recording.

The simplest method for comparing an incoming audio signature (which could be from a file on the Internet, a recording of a radio or Internet radio broadcast, a recording from a cell phone, etc) to a database of reference signatures for the purpose of identification is to simply compare the incoming signature to every element of the database. However, since it may not be known where the reference signatures might have occurred inside the incoming signature, this comparison must be done at many time locations within the incoming signature. Each individual signature-to-signature comparison at each point in time may also be done in a "brute-force" manner using techniques known in the art; essentially computing the full Euclidean distance between the entire signatures' feature vectors. A match can then be declared when one of these comparisons yields a score or distance that is above or below some threshold, respectively.

However, when an audio signature or fingerprint contains a large number of features such a brute-force search becomes too expensive computationally for real-world databases which typically have several hundred thousand to several million signatures.

Many researchers have worked on methods for multi-dimensional indexing, although the greatest effort has gone into geographical (2-dimensional) or spatial (3-dimensional) data. Typically, all of these methods order the elements of the database based on their proximity to each other.

For example, the elements of the database can be clustered into hyper-spheres or hyper-rectangles, or the space can be organized into a tree form by using partitioning planes. However, when the number of dimensions is large (on the order of 15 or more), it can be shown mathematically that more-or-less uniformly distributed points in the space all become approximately equidistant from each other. Thus, it becomes impossible to cluster the data in

a meaningful way, and comparisons can become both lengthy and inaccurate.

Hence, there exists a need to provide a means for data comparison which overcomes the disadvantages of the prior art.

BRIEF DESCRIPTION OF THE INVENTION

A method and apparatus for identifying an unknown work is disclosed. In one aspect, a method may includes the acts of providing a reference database having a reduced dimensionality containing signatures of sampled works; receiving a sampled work; producing a signature from the work; and reducing the dimensionality of the signature.

BRIEF DESCRIPTION OF THE DRAWING FIGURES

Figure 1A is a flowchart of a method according to the present invention.

Figure 1B is a flowchart of another method according to the present invention.

Figure 2 is a diagram of a system suitable for use with the present invention.

Figure 3 is a diagram of segmenting according to the present invention.

Figure 4 is a detailed diagram of segmenting according to the present invention showing hop size.

Figure 5 is a graphical flowchart showing the creating of a segment feature vector according to the present invention.

Figure 6 is a diagram of a signature according to the present invention.

Figure 7A is a flowchart of a method for preparing a reference database according to the present invention.

Figure 7B is a flowchart of method for identifying an unknown work according to the present invention.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Persons of ordinary skill in the art will realize that the following description of the present invention is illustrative only and not in any way limiting. Other embodiments of the invention will readily suggest themselves to such skilled persons having the benefit of this disclosure.

It is contemplated that the present invention may be embodied in various computer and machine-readable data structures. Furthermore, it is contemplated that data structures embodying the present invention will be transmitted across computer and machine-readable

media, and through communications systems by use of standard protocols such as those used to enable the Internet and other computer networking standards.

The invention further relates to machine-readable media on which are stored embodiments of the present invention. It is contemplated that any media suitable for storing instructions related to the present invention is within the scope of the present invention. By way of example, such media may take the form of magnetic, optical, or semiconductor media.

The present invention may be described through the use of flowcharts. Often, a single instance of an embodiment of the present invention will be shown. As is appreciated by those of ordinary skill in the art, however, the protocols, processes, and procedures described herein may be repeated continuously or as often as necessary to satisfy the needs described herein. Accordingly, the representation of the present invention through the use of flowcharts should not be used to limit the scope of the present invention.

The present invention may also be described through the use of web pages in which embodiments of the present invention may be viewed and manipulated. It is contemplated that such web pages may be programmed with web page creation programs using languages standard in the art such as HTML or XML. It is also contemplated that the web pages described herein may be viewed and manipulated with web browsers running on operating systems standard in the art, such as the Microsoft Windows® and Macintosh® versions of Internet Explorer® and Netscape®. Furthermore, it is contemplated that the functions performed by the various web pages described herein may be implemented through the use of standard programming languages such a Java® or similar languages.

The present invention will first be described in general overview. Then, each element will be described in further detail below.

Referring now to Figure 1A, a flowchart is shown which provides a general overview of the present invention as related to the preparation of a database of reference signatures. Two overall acts are performed to prepare a reference database in accordance with the present invention: in act 100, the present invention reduces the dimensionality of reference signatures; and the reference database is indexed in act 102.

Referring now to Figure 1B, a flowchart is shown which provides a general overview of the present invention as related to the identification of an unknown signature in accordance with the present invention. In act 104, a sampled work is received. In act 106, the present invention reduces the dimensionality of the received work. In act 108, the

present invention determines initial candidates. In act 110, the present invention searches for the best candidate.

Prior to presenting a detailed overview of each act of FIGS. 1A and 1B, some background will first be presented.

Structural embodiment of the present invention

Referring now to Figure 2, a diagram of a system suitable for use with the present invention is shown. FIG. 2 includes a client system 200. It is contemplated that client system 200 may comprise a personal computer 202 including hardware and software standard in the art to run an operating system such as Microsoft Windows ®, MAC OS ® Palm OS, UNIX, or other operating systems standard in the art. Client system 200 may further include a database 204 for storing and retrieving embodiments of the present invention. It is contemplated that database 204 may comprise hardware and software standard in the art and may be operatively coupled to PC 202. Database 204 may also be used to store and retrieve the works and segments utilized by the present invention.

Client system 200 may further include an audio/video (A/V) input device 208. A/V device 208 is operatively coupled to PC 202 and is configured to provide works to the present invention which may be stored in traditional audio or video formats. It is contemplated that A/V device 208 may comprise hardware and software standard in the art configured to receive and sample audio works (including video containing audio information), and provide the sampled works to the present invention as digital audio files. Typically, the A/V input device 208 would supply raw audio samples in a format such as 16-bit stereo PCM format. A/V input device 208 provides an example of means for receiving a sampled work.

It is contemplated that sampled works may be obtained over the Internet, also. Typically, streaming media over the Internet is provided by a provider, such as provider 218 of FIG. 2. Provider 218 includes a streaming application server 220, configured to retrieve works from database 222 and stream the works in a formats standard in the art, such as Real®, Windows Media®, or QuickTime®. The server then provides the streamed works to a web server 224, which then provides the streamed work to the Internet 214 through a gateway 216. Internet 214 may be any packet-based network standard in the art, such as IP, Frame Relay, or ATM.

To reach the provider 218, the present invention may utilize a cable or DSL head end

212 standard in the art operatively, which is coupled to a cable modem or DSL modem 210 which is in turn coupled to the system's network 206. The network 206 may be any network standard in the art, such as a LAN provided by a PC 202 configured to run software standard in the art.

It is contemplated that the sampled work received by system 200 may contain audio information from a variety of sources known in the art, including, without limitation, radio, the audio portion of a television broadcast, Internet radio, the audio portion of an Internet video program or channel, streaming audio from a network audio server, audio delivered to personal digital assistants over cellular or wireless communication systems, or cable and satellite broadcasts.

Additionally, it is contemplated that the present invention may be configured to receive and compare segments coming from a variety of sources either stored or in real-time. For example, it is contemplated that the present invention may compare a real-time streaming work coming from streaming server 218 or A/V device 208 with a reference segment stored in database 204.

Segmenting background

It is contemplated that a wide variety of sampled works may be utilized in the present invention. However, the inventors have found the present invention especially useful with segmented works. An overview of a segmented work will now be provided.

Figure 3 shows a diagram showing the segmenting of a work according to the present invention. FIG. 3 includes audio information 300 displayed along a time axis 302. FIG. 3 further includes a plurality of segments 304, 306, and 308 taken of audio information 300 over some segment size T.

In an exemplary non-limiting embodiment of the present invention, instantaneous values of a variety of acoustic features are computed at a low level, preferably about 100 times a second. In particular, 10 MFCCs (cepstral coefficients) are computed. It is contemplated that any number of MFCCs may be computed. Preferably, 5-20 MFCCs are computed, however, as many as 30 MFCCs may be computed, depending on the need for accuracy versus speed.

Segment-level features are disclosed US Patent #5,918,223 to Blum, et al., which is assigned to the assignee of the current disclosure and incorporated by reference as though fully set forth herein. In an exemplary non-limiting embodiment of the present invention,

the segment-level acoustical features comprise statistical measures as disclosed in the '223 patent of low-level features calculated over the length of each segment. The data structure may store other bookkeeping information as well (segment size, hop size, item ID, UPC, etc).

As can be seen by inspection of FIG. 3, the segments 304, 306, and 308 may overlap in time. This amount of overlap may be represented by measuring the time between the center point of adjacent segments. This amount of time is referred to herein as the hop size of the segments, and is so designated in FIG. 3. By way of example, if the segment length T of a given segment is one second, and adjacent segments overlap by 50%, the hop size would be 0.5 second.

The hop size may be set during the development of the software. Additionally, the hop sizes of the reference database and the real-time signatures may be predetermined to facilitate compatibility. For example, the reference signatures in the reference database may be precomputed with a fixed hop and segment size, and thus the client applications should conform to this segment size and have a hop size which integrally divides the reference signature hop size. It is contemplated that one may experiment with a variety of segment sizes in order to balance the tradeoff of accuracy with speed of computation for a given application.

The inventors have found that by carefully choosing the hop size of the segments, the accuracy of the identification process may be significantly increased. Additionally, the inventors have found that the accuracy of the identification process may be increased if the hop size of reference segments and the hop size of segments obtained in real-time are each chosen independently. The importance of the hop size of segments may be illustrated by examining the process for segmenting pre-recorded works and real-time works separately.

Reference signatures

Prior to attempting to identify a given work, a reference database of signatures must be created. When building a reference database, a segment length having a period of less than three seconds is preferred. In an exemplary non-limiting embodiment of the present invention, the segment lengths have a period ranging from 0.5 seconds to 3 seconds. For a reference database, the inventors have found that a hop size of approximately 50% to 100% of the segment size is preferred.

It is contemplated that the reference signatures may be stored on a database such as

database 204 as described above. Database 204 and the discussion herein provide an example of means for providing a plurality of reference signatures each having a segment size and a hop size.

Unknown signatures

The choice of the hop size is important for the signatures of the audio to be identified, hereafter referred to as "unknown audio."

Figure 4 shows a detailed diagram of the segmentation of unknown audio according to the present invention. FIG. 4 includes unknown audio information 400 displayed along a time axis 402. FIG. 4 further includes segments 404 and 406 taken of audio information 400 over some segment length T. In an exemplary non-limiting embodiment of the present invention, the segment length of unknown audio segments is chosen to range from 0.5 to 3 seconds.

As can be seen by inspection of FIG. 4, the hop size of unknown audio segments is chosen to be smaller than that of reference segments. In an exemplary non-limiting embodiment of the present invention, the hop size of unknown audio segments is less than 50% of the segment size. In yet another exemplary non-limiting embodiment of the present invention, the unknown audio-hop size may be 0.1 seconds.

The inventors have found such a small hop size advantageous for the following reasons. The ultimate purpose of generating unknown audio segments is to analyze and compare them with the reference segments in the database to look for matches. The inventors have found at least two major reasons why an unknown audio recording would not match its counterpart in the database. One is that the broadcast channel does not produce a perfect copy of the original. For example, the work may be edited or processed or the announcer may talk over part of the work. The other reason is that larger segment boundaries may not line up in time with the original segment boundaries of the target recordings.

The inventors have found that by choosing a smaller hop size, some of the segments will ultimately have time boundaries that line up with the original segments, notwithstanding the problems listed above. The segments that line up with a "clean" segment of the work may then be used to make an accurate comparison while those that do not so line up may be ignored. The inventors have found that a hop size of 0.1 seconds seems to be the maximum that would solve this time shifting problem.

As mentioned above, once a work has been segmented, the individual segments are then analyzed to produce a segment feature vector. Figure 5 is a diagram showing an overview of how the segment feature vectors may be created using the methods described in US Patent #5,918,223 to Blum, et al. It is contemplated that a variety of analysis methods may be useful in the present invention, and many different features may be used to make up the feature vector. The inventors have found that the pitch, brightness, bandwidth, and loudness features of the '223 patent to be useful in the present invention. Additionally, spectral features may be used analyzed, such as the energy in various spectral bands. The inventors have found that the cepstral features (MFCCs) are very robust (more invariant) given the distortions typically introduced during broadcast, such as EQ, multi-band compression/limiting, and audio data compression techniques such as MP3 encoding/decoding, etc.

In act 500, the audio segment is sampled to produce a segment. In act 502, the sampled segment is then analyzed using Fourier Transform techniques to transform the signal into the frequency domain. In act 504, mel frequency filters are applied to the transformed signal to extract the significant audible characteristics of the spectrum. In act 506, a Discrete Cosine Transform is applied which converts the signal into mel frequency cepstral coefficients (MFCCs). Finally, in act 508, the MFCCs are then averaged over a predetermined period. In an exemplary non-limiting embodiment of the present invention, this period is approximately one second. Additionally, other characteristics may be computed at this time, such as brightness or loudness. A segment feature vector is then produced which contains a list containing at least the 10 MFCCs corresponding average.

The disclosure of FIGS. 3, 4, and 5 provide examples of means for creating a signature of a sampled work having a segment size and a hop size.

Figure 6 is a diagram showing a complete signature 600 according to the present invention. Signature 600 includes a plurality of segment feature vectors 1 through n generated as shown and described above. Signature 600 may also include an identification portion containing a unique ID. It is contemplated that the identification portion may contain a unique identifier provided by the RIAA (Recording Industry Association of America) or some other audio authority or cataloging agency. The identification portion may also contain information such as the UPC (Universal Product Code) of the various products that contain the audio corresponding to this signature. Additionally, it is contemplated that the signature 600 may also contain information pertaining to the

characteristics of the file itself, such as the hop size, segment size, number of segments, etc., which may be useful for storing and indexing.

Signature 600 may then be stored in a database and used for comparisons.

The following computer code in the C programming language provides an example of a database structure in memory according to the present invention:

```
typedef struct
{
        float   hopSize;                /* hop size */
        float   segmentSize;            /* segment size */
        MFSignature* signatures;        /* array of signatures */
} MFDatabase;
```

The following provides an example of the structure of a segment according to the present invention:

```
typedef struct
{
        char*   id;                     /* unique ID for this audio clip */
        long    numSegments;/* number of segments */
        float*  features;               /* feature array */
        long    size;                   /* size of per-segment feature vector */
        float   hopSize;
        float   segmentSize;
} MFSignature;
```

The discussion of FIG. 6 provides an example of means for storing segments and signatures according to the present invention.

A more detailed description of the operation of the present invention will now be provided.

Referring now to Figure 7A, a flowchart showing one aspect of a method according to the present invention is presented.

Reference database preparation

Prior to the identification of an unknown sample, a database of reference signatures is prepared in accordance with the present invention.

In an exemplary non-limiting embodiment of the present invention, a reference signature may comprise an audio signature derived from a segmentation of the original audio work as described above. In a presently preferred embodiment, reference signatures have 20 non-overlapping segments, where each segment is one second in duration, with one-second spacing from center to center, as described above. Each of these segments is represented by 10 Mel filtered cepstral coefficients (MFCCs), resulting in a feature vector of 200 dimensions. Since indexing a vector space of this dimensionality is not practical, the number of dimensions used for the initial search for possible candidates is reduced according to the present invention.

Reducing the dimensionality

Figure 7A is a flowchart of dimension reduction according to the present invention. The number of dimensions used for the initial search for possible candidates is reduced, resulting in what the inventors refer to as a subspace. By having the present invention search a subspace at the outset, the efficiency of the search may be greatly increased.

Referring now to FIG. 7A, the present invention accomplishes two tasks to develop this subspace: (1) the present invention uses less than the total number of segments in the reference signatures in act 701; and (2) the present invention performs a principal components analysis to reduce the dimensionality in act 703.

Using less segments to perform an initial search

The inventors empirically have found that using data from two consecutive segments (i.e., a two-second portion of the signature) to search for approximately 500 candidates is a good tradeoff between computation complexity and accuracy. The number of candidates can be altered for different applications where either speed or accuracy is more or less important.

For example, the present invention may be configured to extract a predetermined percentage of candidates. In an exemplary non-limiting embodiment of the present invention, a list of candidates may comprise 2% of the size of the reference signature database when using 2 segments for the initial search. In another exemplary non-limiting

embodiment of the present invention, a list of candidates may be those reference signatures whose distances based on the initial 2-segment search are below a certain threshold.

As will be appreciated by those of ordinary skill in the art, the dimension reduction of the present invention may be used to perform initial search using fewer segments for data other than MFCC-based feature vectors. It is contemplated that any feature-based vector set may be used in the present invention.

Furthermore, the segments used in the initial search do not have to be the same size as the segments used for the final search. Since it may be better to use as few dimensions as possible in the initial search for candidates, a smaller segment size is advantageous here. The full segment size can then be used in the final search. In an exemplary non-limiting embodiment of the current invention, the initial search may use the higher-order MFCCs (since these are the most robust) - this is a simple way to reduce the dimensionality.

In the next section, we will discuss another, more sophisticated, method for reducing the segment size for the initial candidate search.

Perform alternate encoding

The second step is to use an alternate encoding of the MFCC data which has the same information but with fewer features.

To accomplish this, the present invention first performs an eigenanalysis of N candidates to determine the principal components of the MFCCs for our typical audio data. In an exemplary non-limiting embodiment of the present invention, the present invention examines 25,000 audio signatures of 20 segments each – each taken from a different recording, which gives provides 500,000 sets of MFCCs. The inventors have found that this is enough to be a good statistical sample of the feature vectors.

As is appreciated by those of ordinary skill in the art, the number examined in the present invention may be adjusted to provide a good statistical sample of different kinds of music. For example, 100 or a1000 segments may be satisfactory.

Next, a Karhunen-Loève transformation is derived. Each set of 10 MFCCs becomes a column of a matrix A. We then compute A^TA and find the 10 eigenvalues and eigenvectors of this matrix. Sorting the eigenvectors by eigenvalue (largest eigenvalue first) results in a list of orthogonal basis vectors that are the principal components of the segment data. For a database of typical music recordings, 95% of the information in the MFCCs is contained in the first 7 components of this new basis.

As is known by those having ordinary skill in the art, the Karhunen-Loève transformation is represented by the matrix that has the all 10 of the above eigenvectors as its rows. This transformation is applied to all the segments of all the reference signatures in the database as well as to all the segments of any signatures that are to be identified. This allows approximate distances to be computed by using the first few components of the transformed segment MFCC vectors for a small tradeoff in accuracy. Most importantly, it reduces the initial search dimension to 14 (7 components times 2 segments), which can be indexed with reasonable efficiency.

As will be appreciated by those of ordinary skill in the art, dimension reduction according to the present invention may be utilized to examine subspaces for feature sets other than MFCCs. The dimension reduction of the present invention may be applied to any set of features since such sets comprise vectors of floating point numbers. For example, given a feature vector comprising spectral coefficients and loudness, one could still apply the KL-process of the present invention to yield a smaller and more easily searched feature vector.

Furthermore, the transform of the present invention may be applied to each segment separately. For example, prior art identification methods may use a single 30-second segment of sound over which they compute an average feature vector. Of course, the accuracy of such methods are much lower, but the process of the present invention may work for such features as well. Moreover, such prior art methods may be used as an initial search.

The dimension reduction aspect of the present invention provides significant efficiency gains over prior art methods. For example, in a "brute force" method, the signature of the incoming sampled work is tested against every reference signature in the database. This is time-consuming because the comparison of any two signatures is a 200-dimensional comparison and because there are a lot of reference signatures in the database. Either alone are not unsatisfactory, but both together takes a long time. The present invention solves the first problem by searching only a subspace, i.e., using less than all 200 dimensions in the comparison.

In addition to the raw speedup given by searching a subspace, the reduced dimensionality also allows one to practically index the database of reference signatures. As mentioned above, it is impractical to index a 200-dimensional database, but 14 is practical.

The present disclosure thus provides for several manners in which the dimensionality may be reduced:

(1) searching for the top N candidates over a subspace;

(2) searching for the top N candidates using less than the total number of segments from the reference signature;

(3) searching for the top N candidates by projecting the reference signatures and signature of the work to be identified onto a subspace; and

(4) searching for the top N candidates by projecting the reference signatures and signature of the work to be identified onto a subspace, where the subspace is determined by a Karhunen- Loève transformation.

The preparation of the reference database may occur at any time. For example, the results of the preparation may occur each time the server is started up. Additionally, the results could be saved and reused from then on, or the results may be prepared once and used over again. It may need to be recomputed whenever a new reference signature is added to the database.

Computing the index

The present invention may also compute an index of the reference signatures. As is appreciated by those having ordinary skill in the art, many indexing strategies are available for use in the present invention. Examples include the k-d tree, the SS-tree, the R-tree, the SR-tree, and so on. Any look-up method known in the art may be used in the present disclosure. Common to all indexing strategies is that the multidimensional space is broken into a hierarchy of regions which are then structured into a tree. As one progress down the tree during the search process, the regions become smaller and have fewer elements. All of these trees have tradeoffs that affect the performance under different conditions, e.g., whether the entire tree fits into memory, whether the data is highly clustered, and so on.

In an exemplary non-limiting embodiment of the present invention, a binary k-d tree indexing method is utilized. This is a technique well-known in the art, but a brief overview is given here. At the top level, the method looks to see which dimension has the greatest extent, and generates a hyperplane perpendicular to this dimension that splits the data into two regions at its median. This yields two subspaces on either side of the plane. This process is continued by recursion on the data in each of these subspaces until each of the subspaces has a single element.

After the reference database has been prepared, the present invention may be used to identify an unknown work. Such a process will now be shown and described.

Identification of an unknown work

Referring now to FIG. 7B, a flowchart of a method for identifying an unknown work is shown. In act 700, the present invention receives a sampled work. In act 702, the present invention determines a set of initial candidates. Finally, in act 704, the present invention determines the best candidate. Each act will now be described in more detail.

Receiving a sampled work

Beginning with act 700, a sampled work is provided to the present invention. It is contemplated that the work will be provided to the present invention as a digital audio stream. It should be understood that if the audio is in analog form, it may be digitized in any manner standard in the art.

Indexed lookup

In act 702, the present invention determines the initial candidates. In a preferred embodiment, the present invention uses the index created above to perform an indexed candidate search.

An index created in accordance with the present invention may used to do the N nearest neighbor search required to find the initial candidates.

Candidate search

Once a set of N nearest neighbors is determined, the closest candidate may then be determined in act 704. In an exemplary non-limiting embodiment of the present invention, a brute-force search method may be used to determine which candidate is the closest to the target signature. In another preferred embodiment, the present invention may compare the distance of this best candidate to a predetermined threshold to determine whether there is a match.

There are a number of techniques that may be applied to the candidate search stage which make it much faster. In one aspect, these techniques may be used in a straightforward brute-force search that did not make use of any of the steps previously described above. That is, one could do a brute-force search directly on the reference signature database

without going through the index search of step 702, for example. Since there is some overhead in doing step 702, direct brute-force search may be faster for some applications, especially those that need only a small reference database, e.g., generating a playlist for a radio station that plays music from a small set of possibilities.

Speedups of brute-force search

Any reference signature that is close to the real-time signature has to be reasonably close to it for every segment in the signature. Therefore, in one aspect, several intermediate thresholds are tested as the distance is computed and the computation is exited if any of these thresholds are exceeded. In a further aspect, each single segment-to-segment distance is computed as the sum of the squared differences of the MFCCs for the two corresponding segments. Given the current computation of the MFCCs, average segment-to-segment distances for matches are about approximately 2.0. In an exemplary non-limiting embodiment of the invention, we exit the computation and set the distance to infinity if any single segment-to-segment distance is greater than 20. In further aspects, the computation is exited if any two segment-to-segment distances are greater than 15, or if any four segment-to-segment distances are greater than 10. It should be clear to anyone skilled in the art that other thresholds for other combinations of intermediate distances could easily be implemented and set using empirical tests.

Since any match will also be close to a match at a small time-offset, we may initially compute the distances at multiples of the hop size. If any of these distances are below a certain threshold, we compute the distances for hops near it. In an exemplary non-limiting embodiment of the invention, we compute distances for every third hop. If the distance is below 8.0, we compute the distances for the neighboring hops. It should be clear to anyone skilled in the art that other thresholds for other hop-skippings could easily be implemented and set using simple empirical tests.

While embodiments and applications of this invention have been shown and described, it would be apparent to those skilled in the art that many more modifications than mentioned above are possible without departing from the inventive concepts herein. For example, the teachings of the present disclosure may be used to identify a variety of sampled works, including, but not limited to, images, video and general time-based media. The invention, therefore, is not to be restricted except in the spirit of the appended claims.

What is claimed is:

1. A method for identifying an unknown work comprising:

providing a reference database having a reduced dimensionality containing signatures of sampled works;

receiving a sampled work;

producing a signature from said work; and

reducing the dimensionality of said signature.

2. The method claim 1, further including the act of comparing said signature of said received work to said reference database.

3. The method of claim 2, further including the act of generating a list of candidates.

4. The method of claim 3, further including the act of comparing the non-reduced signature to the non-reduced candidate signatures to obtain a result.

5. The method of claim 1, wherein said signature includes a plurality of MFCCs calculated for each said segment.

6. The method of claim 5, wherein said sampled work includes a signature calculated by using a plurality of acoustical features from the group consisting of at least one of loudness, pitch, brightness, bandwidth, spectrum and MFCC coefficients.

7. The method of claim 1, wherein said act of reducing the dimensionality is performed by using an indexing strategy chosen from the group consisting of: the k-d tree, the SS-tree, the R-tree, and the SR-tree.

8. The method of claim 1, wherein said sampled work signature comprises a plurality of segments and an identification portion.

9. The method of claim 1, wherein said sampled work comprises a segment size of approximately 0.5 to 3 seconds.

10. The method of claim 1, wherein said sampled work comprises a segment size of approximately 1 second.

11. The method of claim 1, wherein said sampled work comprises a hop size of less than 50% of the segment size.

12. The method of claim 1, wherein said sampled work comprises a hop size of approximately 0.1 seconds.

13. The method of claim 1, wherein said signature contains averages taken from the group consisting of: loudness, pitch, brightness, bandwidth, spectral features, and

cepstral coefficients.

14. The method of claim 1, wherein said act of reducing the dimensionality is performed by projecting the features onto a Karhunen-Loeve basis.

15. The method of claim 1, wherein said act of reducing the dimensionality is performed by brute force comparisons.

16. The method of claim 15, wherein 500 candidates are produced.

17. An apparatus for identifying an unknown work comprising:

means for providing a reference database having a reduced dimensionality containing signatures of sampled works;

means for receiving a sampled work;

means for producing a signature from said work; and

means for reducing the dimensionality of said signature.

18. The apparatus of claim 17, further including means for comparing said signature of said received work to said reference database.

19. The apparatus of claim 18, further including means for generating a list of candidates.

20. The apparatus of claim 19, further including means for comparing the non-reduced signature to the non-reduced candidate signatures to obtain a result.

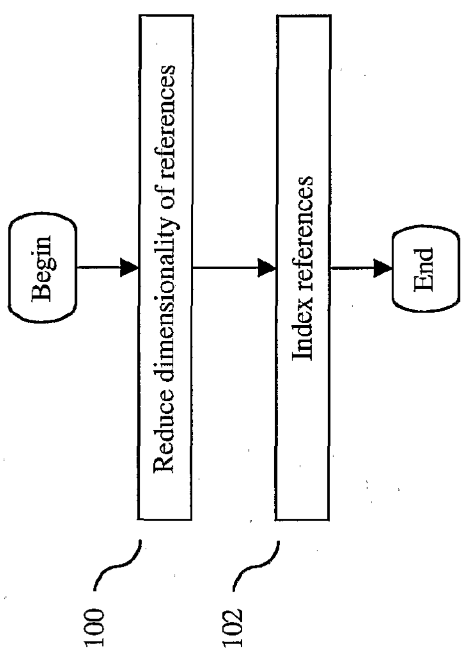


FIG. 1A
Present Invention

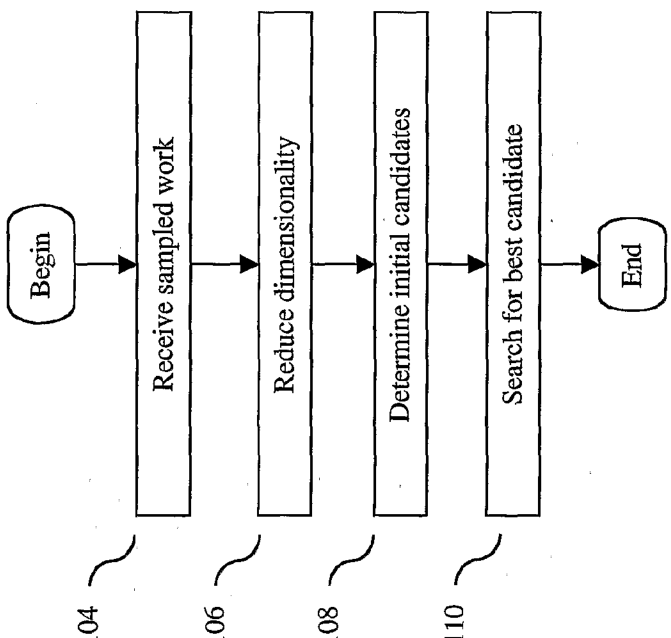


FIG. 1B
Present Invention

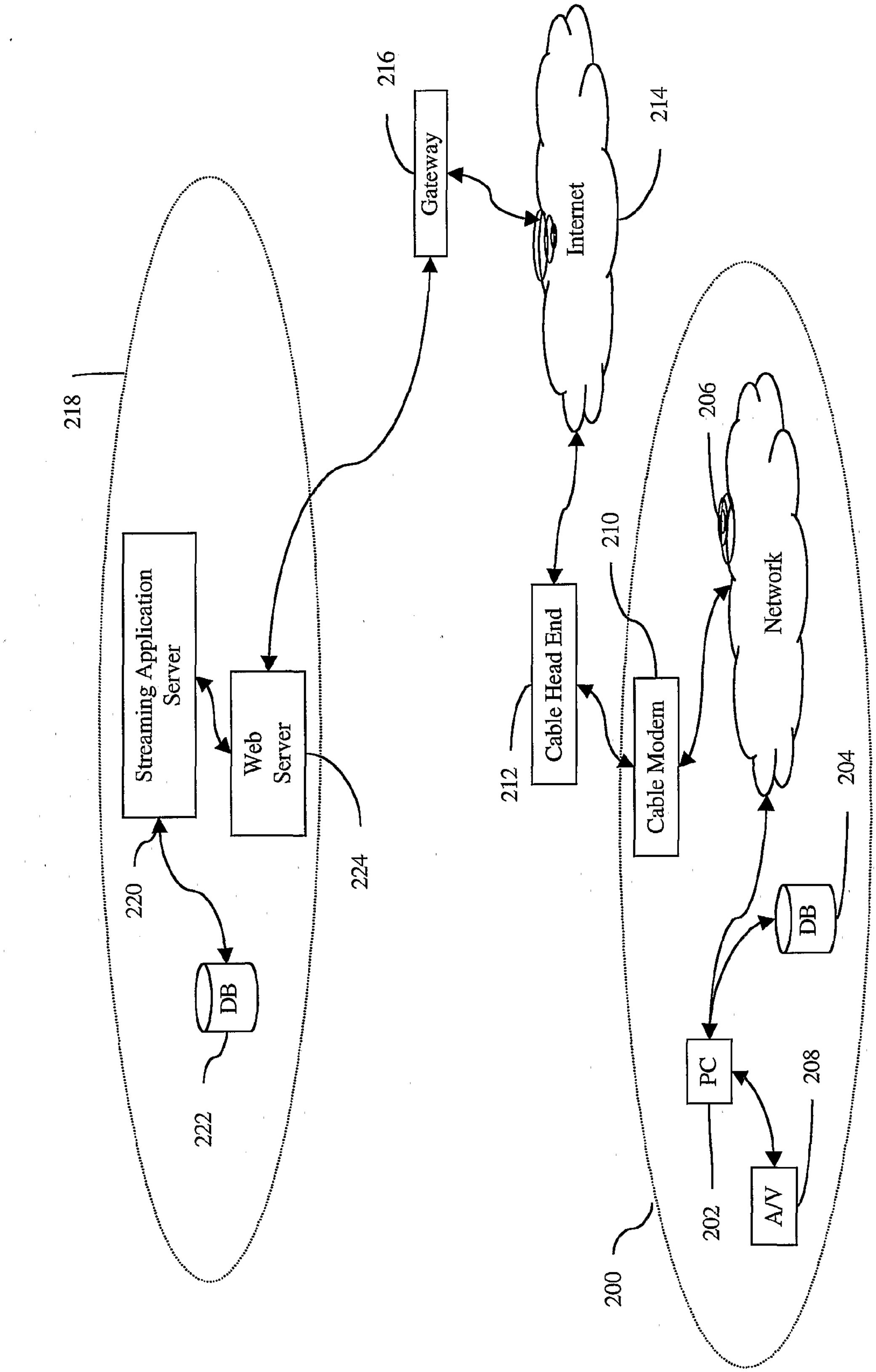


FIG. 2
Present Invention

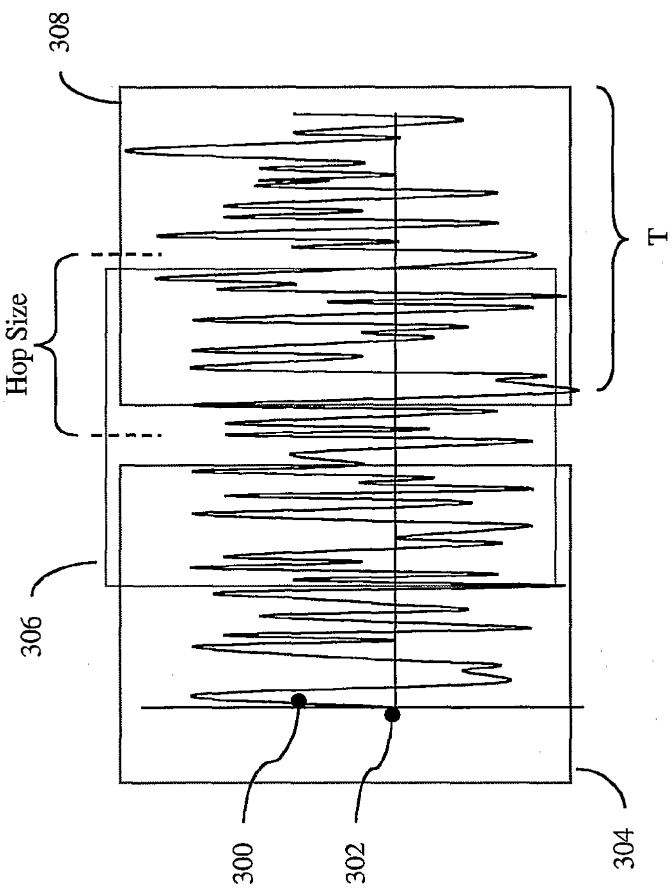


FIG. 3
Present Invention

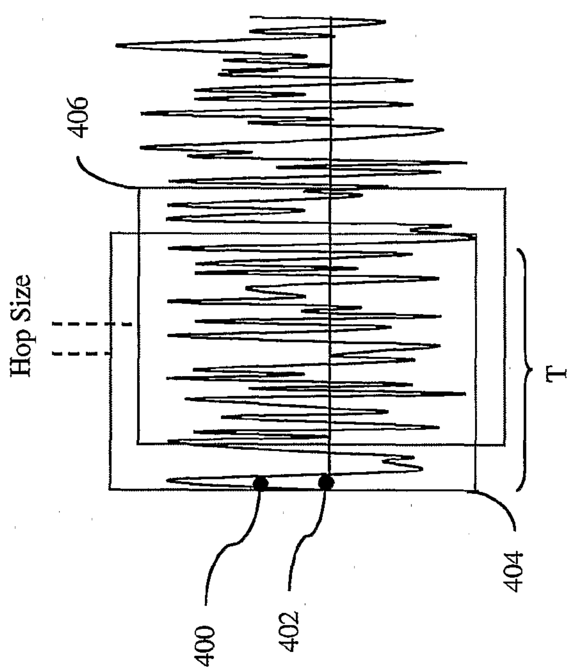


FIG. 4
Present Invention

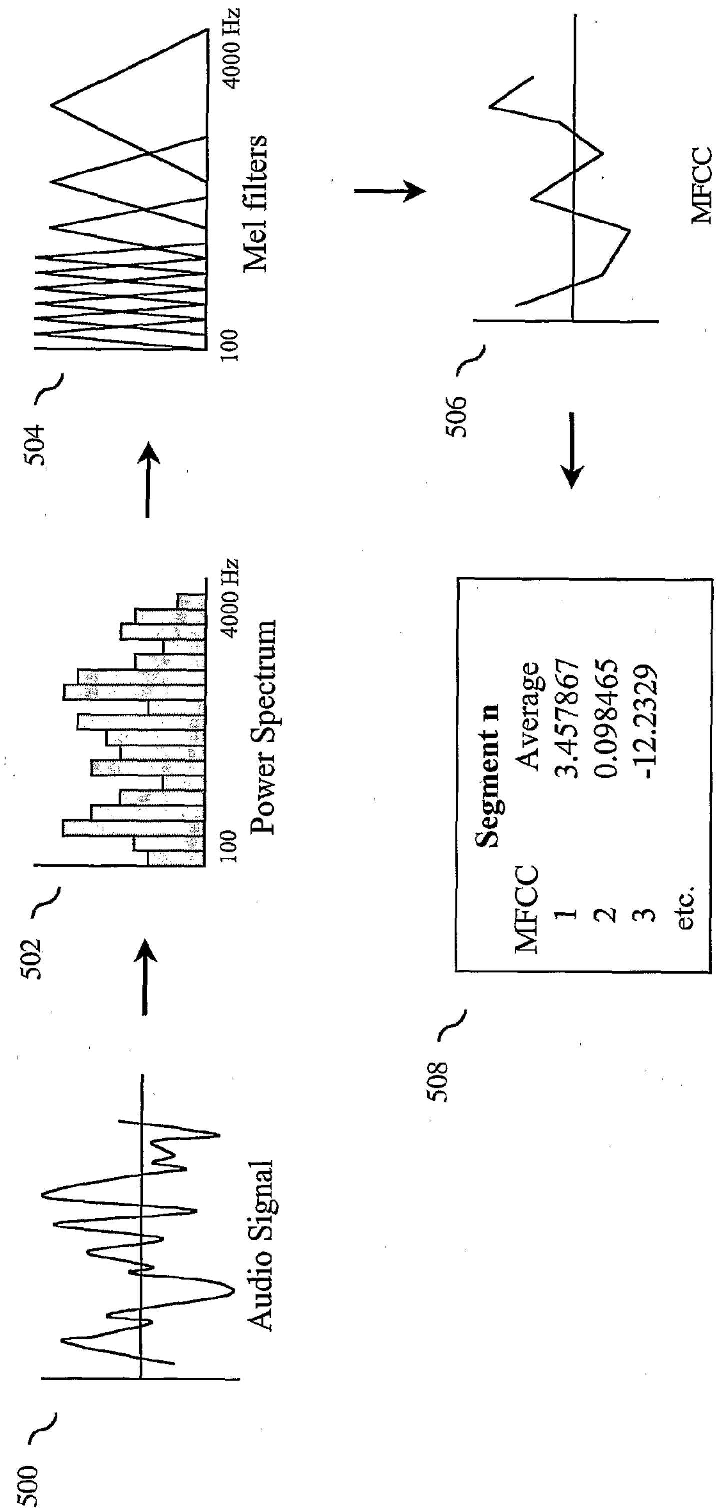


FIG. 5
Present Invention

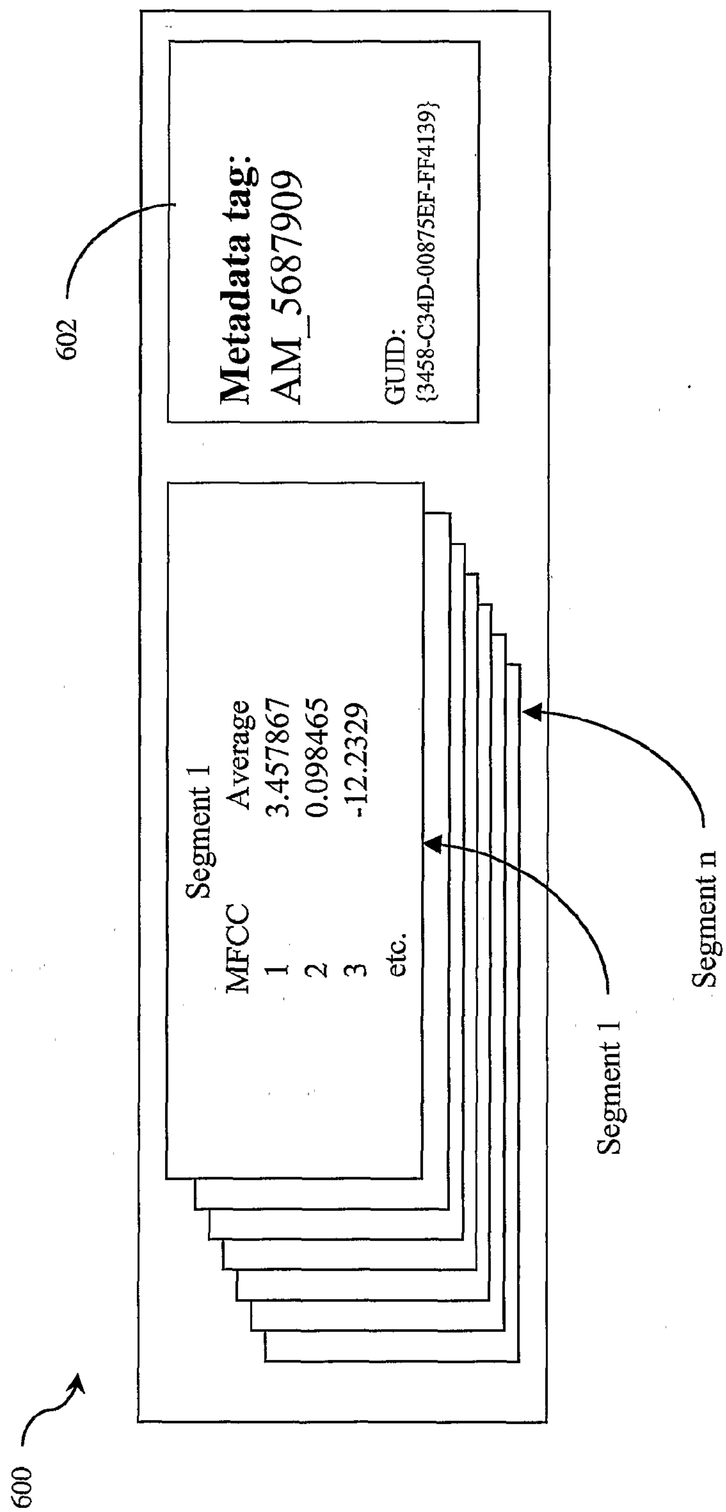


FIG. 6
Present Invention

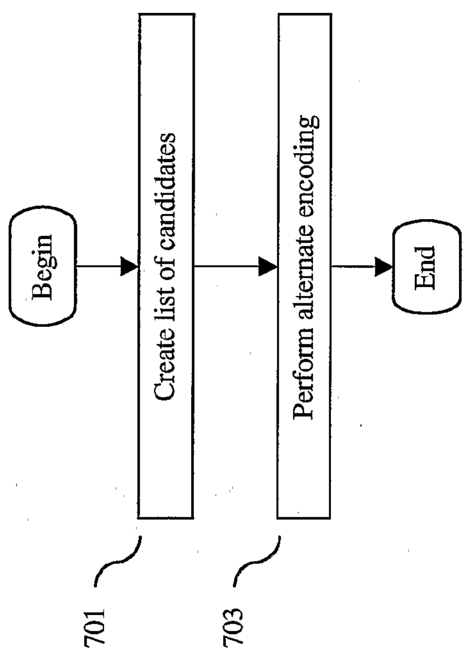


FIG. 7A
Present Invention

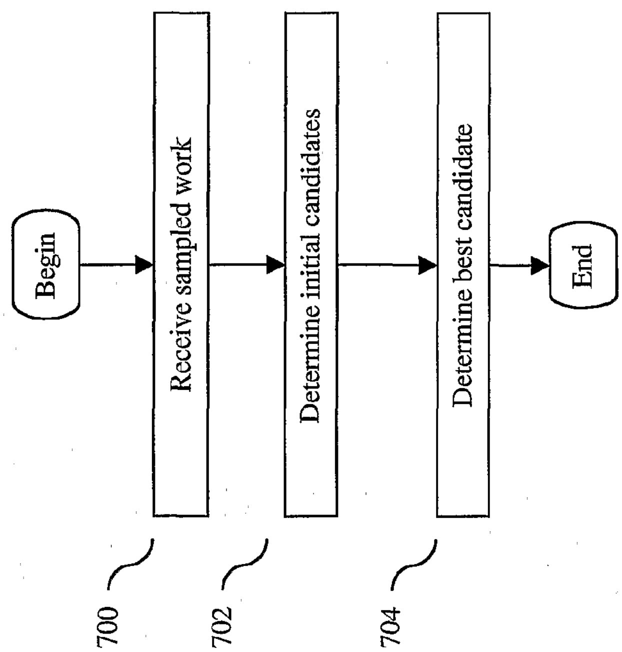


FIG. 7B
Present Invention

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/22460

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G06N 5/02
US CL : 706/45
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 706/45

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EAST, IEEE, ACM, INTERNET

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,710,916 A (BARBARA et al.) 20 January 1998, col. 2, lin. 10-67.	1-20
A	US 6,243,615 B1 (NEWAY et al) 05 June 2001, col. 4, lin. 1-67.	1-20
A, E	US 6,422,061 B1 (SUNSHINE et al.) 23 July 2002, col. 4, lin. 1-67.	1-20

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 13 SEPTEMBER 2002	Date of mailing of the international search report 05 NOV 2002
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer WILBERT L. STARKS, JR. Telephone No. (703) 308-9700